DOCUMENT RESUME

ED 222 054 FL 013 220

AUTHOR Gordon, Peter
TITLE Early Encoding of the Count/Mass Distinction: Semantic or Syntactic?
INSTITUTION Stanford Univ., Calif. Dept. of Linguistics.
PUB DATE 82
NOTE 10p.; In its: Papers and Reports on Child Language Development, Number 21, p71-78.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Child Language; *Classification; *Language Acquisition; *Nouns; Preschool Children; Semantics; Syntax; Verbal Development
IDENTIFIERS Encoding

ABSTRACT

The basis for acquisition of categories in child language was investigated. The early encoding of the distinction between mass and count nouns was examined to determine whether children categorize them on the basis of semantic type or syntactic regularities. An experiment was designed in which semantic and syntactic cues were in competition: objects were inappropriately paired with mass noun frames and substances with count noun frames. A total of 44 children aged 3 to 5 were taught nonsense words in either a count or mass noun syntactic context. To test for categorization, the subjects were required to complete a sentence. Nearly all subjects categorized on the basis of syntactic context. However, the appropriate pairing of semantic and syntactic cues did cause a significant number of the younger subjects to produce inconsistent patterns. Thus, while there is no evidence for a semantic encoding of the mass/count distinction, the results for the younger subjects were inconclusive. (RW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Papers and Reports on Child Language Development, No. 21 (1982)

Stanford Univ. Dept. of Linguistics

Early Encoding of the Count/Mass Distinction: Semantic or Syntactic?

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY Stanford Univ.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED222054

FL013220

Peter Gordon

Department of Psychology, M.I.T.

One of the most prominent questions in the study of child language concerns the extent to which the child's linguistic representations are semantically rather than syntactically based. One aspect of this question involves the early encoding of syntactic categories. That is, does the child acquire categories on the basis of their semantic regularities and thus encode these categories as semantic types such as: thing, action, description etc., or rather, does s/he encode them in terms of their syntactic regularities. In addressing this issue, we shall examine the early encoding of the count/mass distinction, a subcategorization of the noun class.

In a recent paper on the development of syntactic categories, Maratsos & Chalkley (1981) have argued strongly against the view that the child's early categories are semantically based. Instead they propose that the acquisition process involves the correlation of distributional frames, into which insertion of the category is permissable. For example, rather than proposing that VERB is encoded as a semantic notion such as action, the category is said to be defined over a set of linguistic contexts shared by members of that category, but not by members of some other category such as NOUN or ADJECTIVE.

As an example of categorization on a purely formal basis, the acquisition of gender has been the subject of a great deal of study. In particular, Karmiloff-Smith (1979), in a series of experiments in French, employed a word-learning paradigm to examine three factors on which the child might be basing his/her categorization for gender: 1) The formal gender of the preceding article (un vs une); 2) The phonological properties of the final syllable of the noun (e.g., forcienne = feminine; bicron = masculine); 3) The natural gender of the referent (male vs female). What was found was that children used only formal properties in assigning gender to the noun, the natural gender of the referent was never used as a basis for assigning gender in younger children.

There are very clear reasons why the child should not use natural gender as a cue to syntactic gender as the vast majority of nouns do not possess natural gender as a semantic property and it would thus be unreliable as a cue. We thus have a case in which the child is able to acquire a subcategorization on a purely formal basis. What is also pointed out by Maratsos & Chalkley (Ibid) is that this process occurs quite early and with relative ease. This therefore provides strong evidence that there are no a priori constraints to initially only acquire those categories with a clear semantic correspondence.

The question to be addressed here is, what does the child do in acquiring distinctions where a semantic correspondence does exist? That is, will s/he encode such distinctions semantically, or will s/he carry out the analysis on a purely formal basis as with the case of gender? In addressing this issue, the count/mass distinction presents itself as an ideal test case. Unlike gender, the count/mass distinction does have a clear correspondence to a semantic distinction for most concrete nouns. That is, most count nouns refer to objects or things such as cars, tables and books whereas most mass nouns refer to substances or stuff such as water, sand and gold. Formally, the two classes of nouns are distinguished by determiner selection and pluralisation. For example, only count nouns may be preceded by a, another, many, several, each etc. (e.g., Several Cars vs. *Several Water), only mass nouns may be preceded by much (e.g., Much Water vs.*Much Car/s). Also, in most circumstances, mass nouns may only appear in the non-plural form unlike count nouns which may be either singular or plural.

While the semantic characterization of the distinction (i.e., objects vs substances) does allow prediction of the subcategorization of most concrete nouns, there are clear exceptions. For example, certain superordinate terms such as furniture, equipment and silverware refer to collections of objects and yet are categorized as mass nouns syntactically. Similarly, the terms substance, solution and sample (as in blood-sample) all refer to substances yet are categorized as count nouns. Thus, while the syntactic-to-semantic correspondence for the distinction is not arbitary, the actual subcategorization of any particular noun is formally rather than semantically determined. That is to say, a noun is either count or mass not because of its semantic properties but because of its occurence in particular noun-phrase contexts. In other words, the distinction is a truly syntactic one.

However, it remains to be seen if the distinction is syntactic for the young child, or whether instead it is encoded along semantic dimensions. In testing this question, we will examine how the child assigns categorization as [+count] or [+mass] in the acquisition of new nouns. The two approaches to be contrasted here are illustrated in figures 1 and 2. In figure 1 there is a direct mapping of conceptual types onto linguistic categories. Thus, in assigning categorization to a new noun, if the noun refers to an object the child categorizes it as a count noun. If it refers to a substance , s/he categorizes it as a mass noun.

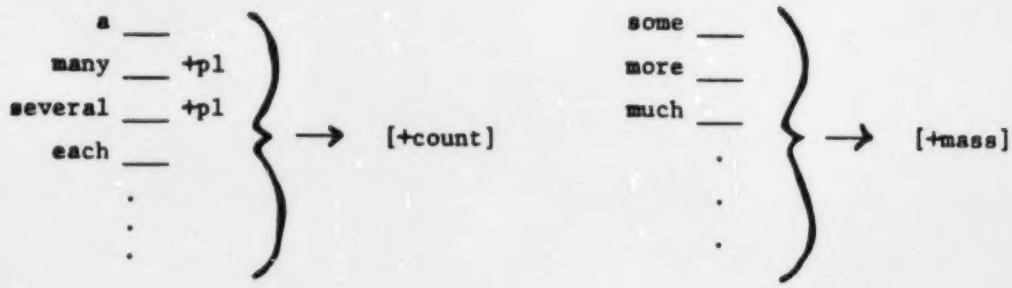
The alternative hypothesis shown in figure 2 proposes that the child assigns categorization on the basis of linguistic context. For example, if the word being learned is 'latt', then if the noun-phrase used is "a latt" then the child categorizes the word as a count noun. If the noun-phrase is "some latt" s/he categorizes it as a mass noun, and similarly for the other relevant frames (see fig 2). Notice that the latter, syntactic hypothesis does not exclude the possibility of semantic coordination, that is, the child may well have correlated the syntactic categories with their corresponding semantic characteristics. What is claimed is that those semantic characteristics do not form the basis for the encoding of the distinction. Thus while Brown (1957) has shown that 3- to 5-year-

Fig. 1 Semantic Encoding.

OBJECT → [+ count]

SUBSTANCE → [+ mass]

Fig. 2 Syntactic Encoding



olds are able to choose objects as referents for nonsense words in count noun contexts (e.g., "This is a latt"), and substances for words in mass noun contexts (e.g., "This is some latt"), this does not demonstrate that the encoding of the distinction is itself semantic in the sense employed here. What we are testing then, is whether the child uses semantic or syntactic cues as the basis for subcategorizing nouns.

In order to test the two competing hypotheses, an experiment was designed in which semantic and syntactic cues were in competition. That is, objects were innappropriately paired with mass noun frames, and substances with count noun frames. Children were taught nonsense words in either a count or a mass noun syntactic context. The referents for these words were either objects or substances, examples of which are shown in figure 3. For example, the child would be shown some of the objects which would be unfamiliar(e.g., electrical components). S/he would then be told: "This is some latt, have you ever seen any latt before? Here's some green latt and here's some blue latt". etc. In the case of the substances, these were presented in four test tubes. The child would be shown one of the tubes and pointing at the liquid inside and shaking it around, the experimenter would say: "This is a latt, have you ever seen a latt before? Here's another latt, this is a green latt and this is another green latt". Pointing to the substances and shaking them were done in order to make clear to the child that it was the substance and not the tube that was being refered to.

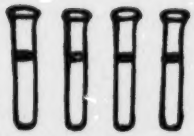
To test for categorization the child was required to complete a sentence. Thus, having been told the name of the object or substance, the child would be told: "Over here we have some more...what?". If the subject completed with a plural form (i.e., more latts) this was taken as indicating that the word had been categorized as a count noun. If a non-plural form was produced (i.e., more latt) this was

taken to indicate categorization as a mass noun.

Fig. 3 Stimulus Items

OBJECTS

SUBSTANCES



With respect to our two hypotheses, we would predict that if the child had encoded the distinction based on semantic properties as in fig 1, then s/he would pluralise when referents were objects but not when they were substances. However, if the encoding of the distinction were on a syntactic basis as in fig 2, then the child should pluralise according to the noun-phrase context regardless of the semantic properties of the referents. Subjects received four items within this condition: two object referents and two substance referents. The pattern of responses over these four items was classified for each subject according to whether the responses indicated categorization on a semantic or syntactic basis.

Fig 4 Classification of Response Patterns.

		some latt + OBJECT		a latt + SUBSTANCE	
(S)	SYNTACTIC	-	-	+	+
($\vec{s}$)	Approaching Syntax	$\begin{bmatrix}-\\-\end{bmatrix}$	$\begin{bmatrix}+\\-\end{bmatrix}$	$\begin{bmatrix}+\\+\end{bmatrix}$	$\begin{bmatrix}+\\-\end{bmatrix}$
(M)	SEMANTIC	+	+	-	-
($\vec{m}$)	Approaching Semantics	$\begin{bmatrix}+\\+\end{bmatrix}$	$\begin{bmatrix}-\\+\end{bmatrix}$	$\begin{bmatrix}-\\-\end{bmatrix}$	$\begin{bmatrix}-\\+\end{bmatrix}$
(Eq)	Equivocal	+	-	+	-
(RB)	Response Bias $\frac{+}{-}$	$\begin{bmatrix}+\\-\end{bmatrix}$	$\begin{bmatrix}+\\-\end{bmatrix}$	$\begin{bmatrix}+\\-\end{bmatrix}$	$\begin{bmatrix}+\\-\end{bmatrix}$

\+ = plural response
\- = non-plural response

The basic patterns are shown in figure 4. Plusses indicate plurals and minuses indicate non-plurals. If the child pluralised only according to syntactic context (i.e., - for "some latt", + for "a latt") then the pattern was classified as 'Syntactic' (S). If three of the four responses were of this type, the pattern was classified as 'Approaching Syntax' ($\hat{S}$). If pluralisation was based on semantic properties (i.e., - for substance, + for object), then the pattern was classified as 'Semantic' (M). 'Approaching Semantics'($\hat{M}$) indicated that three of the four responses were of this type. Less interpretable were the'Response Biases'(always pluralising or never pluralising)(RB+, RB-) and the 'Equivocal' (Eq) patterns in which plurals and non-plurals were equally distributed among count and mass frames. Presumably the (Eq) patterns indicated simple random responding by subjects.

In addition to the "Inappropriate Semantics" condition outlined above, children were also given four items in which syntax and semantics were appropriately paired. That is, objects were referred to as: "a latt" and substances as "some latt". This would allow for a comparison of the two conditions to test whether the appropriate pairing of syntax and semantics would produce different response patterns to when they were inappropriately paired. Also a real word condition was employed in which children were tested with the nouns penny and water under similar presentation conditions. This was essentially a control condition to ensure that the pluralisation test was effective in indicating the categorization of the noun. Also, as was pointed out, the substances were presented in four test-tubes. This was to allow for pluralisation to be plausible for such stimuli, that is one would not expect the child topluralise when the stimulus was either singular or some non-discrete mass. Presenting the water under similar conditions (i.e., in test-tubes) would ensure that the mode of presentation did not bias the child to incorrectly pluralise a mass noun under such conditions.

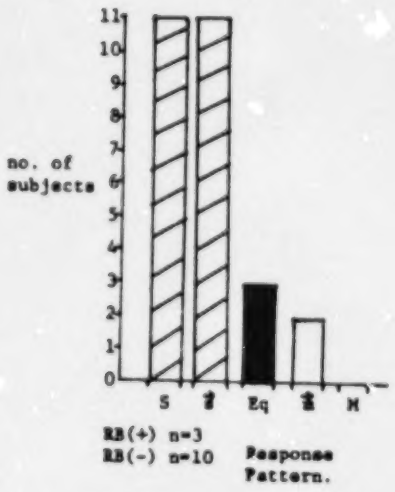
Before initiating the test, subjects were given training in sentence completion and pluralisation. This essentially involved a Berko-type procedure (Berko 1958). Subjects were shown a picture of an object, were told it's name (a nonsense word), were then shown a picture of two of the objects and were told: "Here we have two... what?". Subjects were given feedback and assistance until they produced three consecutive correct responses without assistance. Subjects were eliminated from the analysis if they failed to reach this criterion within fifteen trials.

To summarise the procedure, there were four conditions: 1) Training for sentence completion and pluralisation; 2) The control condition using the words penny and water; 3) Appropriate pairing of semantic and syntactic cues; 4) Inappropriate pairing of semantic and syntactic cues.

A total of 44 subjects were initially tested. Three were eliminated after failing the training phase and one for incorrectly pluralising in the real word condition. Thus the final population consisted of fourty subjects with ages ranging from 3;5 to 5;5, mean age 4;3. For purposes of age-trend analysis, the population was divided into two groups of twenty by age. Group I ranged from 3;5 to 4;1, mean age 3;9. Group II ranged from 4;2 to 5;5, mean age 4:9.

Two main analyses were carried out on the data: 1) an examination of the response patterns obtained in the 'Inappropriate Semantics' (-SEM) condition to ascertain whether categorization was based on semantic or syntactic cues; 2) a comparison of responses for the appropriate- and inappropriate-semantics conditions to determine whether inappropriate pairing affected responses. For the first analysis, subjects responses were classified according to the patterns in fig 4. Overall results are plotted in figure 5. As can be observed from the graph, the vast majority of subjects fell within the syntactic classifications (i.e., S and $\vec{s}$). No subjects consistently responded on the basis of semantic properties (M), and only two approached such a pattern ($\vec{m}$) ($\chi^2 = 42.8$, $p < .001$).

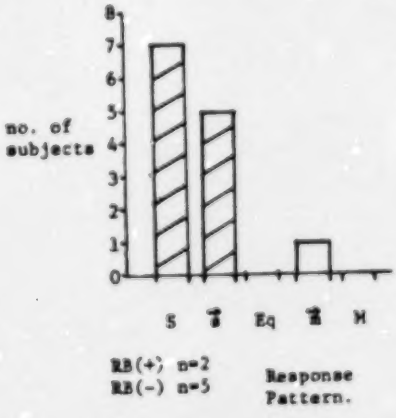
Fig 5 Overall Response Patterns



The patterns for age groups I and II are shown in figures 6 and 7 respectively. Again the patterns for both groups significantly are of the syntactic type with the older subjects being slightly more polarized in this direction (Gp I: $\chi^2 = 7.46$, $p = .03$; Gp II: $\chi^2 = 14.7$, $p < .001$). However, in neither group was there evidence for encoding of the distinction on a semantic basis (i.e., there was not a shift from a semantic to a syntactic basis as a function of age). There is thus clear evidence that children who were able to produce patterns indicating consistent categorizations were nearly all basing these categorizations on the linguistic context and not the semantic properties of the referent.

Fig 6 Response Patterns for Group I

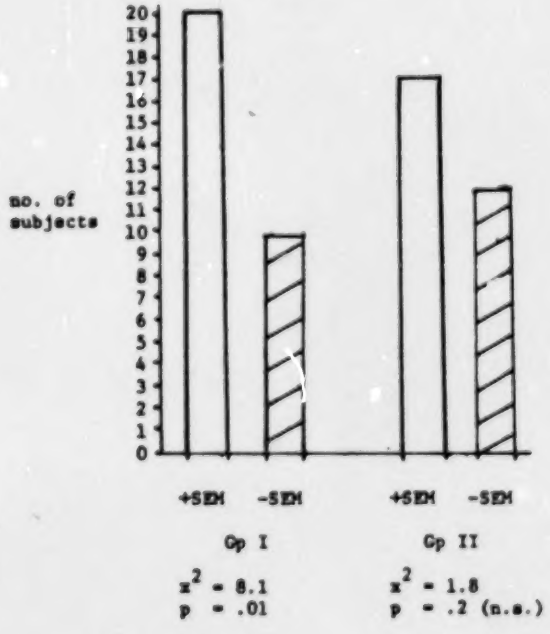
Fig 7 Response Patterns for Group II



It should be noted that a substantial number of subjects did not produce consistent patterns. That is, they either had response biases as shown at the bottom of figs 5 to 7, or responses were equivocal (black bars). One must assume that these subjects were unable to categorize the nouns on any consistent basis.

From the data presented it is clear that the semantic hypothesis is not supported. Children do not categorize nouns on the basis of semantic properties in preference to syntactic properties. However, does this mean that the distinction is essentially no different from gender in being completely formally based for the young child? That is, does it make any difference at all what the semantic properties of the referents are? To address this question a comparison of response patterns was carried out for the 'Inappropriate-Semantics' (-SEM) and the 'Appropriate Semantics' (+SEM) conditions. Syntactic (S) and 'Approaching Syntax' ($\tilde{S}$) patterns were grouped together for the Inappropriate-Semantics condition and compared with the number of equivalent patterns obtained for the Appropriate-Semantics conditions. We are thus comparing the extent to which children produced patterns consistent with the syntactic frame of the noun when semantic properties of referents were either in accord or not in accord with the given subcategorization.

Fig 8. Comparison of S and $\tilde{S}$ Response Patterns for Appropriate-Semantics (+SEM) and Inappropriate-Semantics (-SEM) conditions.



The comparison of these patterns is shown in fig. 8. White bars represent Appropriate-Semantics (+SEM) and striped bars represent Inappropriate-Semantics (-SEM). As can be seen, there was an effect of appropriateness of the semantic properties although this effect was only significant for the younger subjects. What seems to be happening is that while most of the younger subjects were able to produce consistent patterns when there was an appropriate pairing of semantic and syntactic cues, a significant number of them were thrown off when this pairing was inappropriate and either produced response biases or random responses.

It would seem therefore that semantic properties were effective in influencing responses for about half of the younger subjects. Thus, while the semantic hypothesis as stated was not supported, that is subjects did not categorize nouns on the basis of semantic properties, it is also not the case that such properties played no role for the younger subjects. That is, there does appear to be a difference between this case and the case of gender acquisition where it was found that semantic cues were totally ineffective. What exactly the difference is, is not strictly discernable from the results presented here. Being a count noun does imply quantification as a discrete entity, whereas being a mass noun indicates quantification as a non-discrete entity. Thus, while a noun like solution is intrinsically continuous, semantically speaking, a phrase like "several solutions" does imply individuation by either type or measure. Similarly a phrase like "some furniture" implies non-individuation of the set of objects refered to. That is, one is to ignore the actual discreteness of the referent. Thus, while the subcategorization of any particular noun is formally determined by its distributional properties, there are still semantic consequences in terms of whether it is quantified as a discrete or a non-discrete entity (i.e., by the semantics of the actual quantifiers). It is possible that for younger children the actual discreteness of the referent must more closely match the mode of quantification as determined by its subcategorization as a count or a mass noun.

It may well be the case that the present experiment was somewhat artificial in terms of the large number of new words being learned and the rather brief exposure to them. Another study is presently underway in which real words are being tested for subcategorization using a similar pluralisation method. In particular, superordinates such as furniture, silverware and jewelry which are "inappropriate" mass nouns are compared with comparable count noun superordinates such as toy, flower and pet. So far the results suggest that children make very few errors in either case and that errors are no greater for those terms with inappropriate semantic properties with respect to their subcategorizations. Such results, though preliminary, would suggest that in the natural word learning situation with a greater amount of exposure to a variety of contexts, the inappropriateness of semantic properties is less of a hindrence to categorization.

To summarise, we have examined the question of whether children will categorize new nonsense words for the count/mass distinction on the basis of semantic or syntactic properties. When consistent responses were obtained, nearly all subjects categorized on the basis of syntactic context. However, the inappropriate pairing of semantic and syntactic cues did cause a significant number of younger subjects to produce inconsistent patterns that were either biased or equivocal. Thus while there is no evidence for a semantic encoding of the distinction of the type tested here, it would seem that many younger children in this task were not able to subcategorize on a purely formal basis such that the semantic properties of the referents were essentially irrelevant to the process.

I would like to thank the children and staff of the following centers for their cooperation in the study: Newton Jewish Community Center, Weston Jewish Community Center, M.I.T. Daycare, Eastgate. Joseph Hoch and Josette Teuscher ran subjects and Susan Carey, Elissa Newport and Annette Karmiloff-Smith provided advice and comments,to all of whom I am most grateful.

References

Berko, J. (1958) "The Child's Learning of English Morphology" Word, 14, pp150-157.

Brown, R. (1957) "Linguistic Determinism and the Part of Speech" Journal of Abnormal & Social Psychology, 55, pp1-5.

Karmiloff-Smith, A. (1979) "A Functional Approach to Child Language" Cambridge University Press.

Maratsos, M.P. & Chalkley, M.A. (1981) "The Internal Language of Children's Syntax: The Ontogenesis and Representation of Syntactic Categories". In K. Nelson (ed) Children's Language Vol II. New York, Garner Press.